Objective Prediction of Cloud Amount Based on Model Output Statistics

GARY M. CARTER AND HARRY R. GLAHN

Techniques Development Laboratory, National Weather Service, NOAA, Silver Spring, Md. 20910

(Manuscript received 21 July 1976, in revised form 10 September 1976)

ABSTRACT

We have applied the Model Output Statistics (MOS) approach to the prediction of cloudiness. Final guidance warm and cool season forecasting equations were developed by screening forecast fields from the primitive equation and trajectory models. We derived separate equations for each of 233 stations to estimate the probability of clear, scattered, broken and overcast conditions 12 to 48 h in advance. The same predictors were used in all four equations for any given station and projection. In like manner, we also derived a set of early guidance equations for the warm season by screening forecasts from the limited-area fine mesh model. Here, separate equations were developed for 230 stations and projections of 6 to 24 h. Weather parameters from surface reports were also included as potential predictors for the first two forecast projections to provide the latest observed conditions for the early and final guidance systems.

We verified both experimental and operational cloud forecasts made from the final guidance equations for approximately 90 widely distributed test stations. These objective cloud forecasts were compared with subjective National Weather Service local forecasts after transforming the objective probability estimates into categorical form. However, using the category with the highest probability produced too many forecasts of clear and overcast. So we transformed the objective estimates in such a way that the percentage of correct forecasts was still high, but with the restriction that the categorical forecasts were relatively unbiased (i.e., each category of cloud amount was forecast about as often as it occurred). The verification scores showed that both the experimental and operational objective forecasts compared favorably with the subjective forecasts.

1. Introduction

Estimates of cloudiness are of interest to both the aviation industry and the general public. Unfortunately, very little objective cloud forecasting guidance has been available to National Weather Service (NWS) aviation and public weather forecasters. However, since December 1974, objective guidance forecasts of cloud amount (Fig. 1) have been provided twice daily for projections of 12 to 48 h by a system developed by the Techniques Development Laboratory (TDL).

We used a numerical-statistical technique called Model Output Statistics (MOS) to derive cloud forecasting equations for approximately 233 individual stations similar to a system described briefly by Glahn (1974a). Klein and Glahn (1974) have shown how the MOS technique has been applied to the prediction of many other weather elements as well.

Our objective cloud forecasts are in terms of the probability of occurrence of four categories, which correspond roughly to clear, scattered, broken and overcast. We also provide a single "best" category forecast for each station.

2. Development of forecasting equations

The MOS technique consists of determining a statistical relationship between predictand data and variables forecast by one or more numerical models. This requires an extensive collection of forecasts from the relevant numerical model (or models). Such a sample of predictor data from the limited-area fine mesh (LFM) [Howcroft and Desmaris, 1971], primitive equation (PE) [Shuman and Hovermale, 1968] and trajectory (TJ) [Reap, 1972] models, as well as the corresponding observed sky cover predictands, were available in TDL's MOS data collection (Glahn, 1974b).

a. PE-based and TJ-based forecasting equations

Initially, we derived a set of cloud amount forecasting equations by using screening regression to relate cloud amount predictand data to forecasts from the PE and TJ models. Our MOS developmental data sample covered the period October 1969 through March 1974. We separated this data sample into warm seasons of April to September and cool seasons of October to March.

The main predictors screened were measures of moisture at particular levels or integrated through the column, heights and temperatures at constant pressure surfaces, various measures of stability, and U and V wind components, all at differing projection times and levels throughout the atmosphere. The sine and cosine

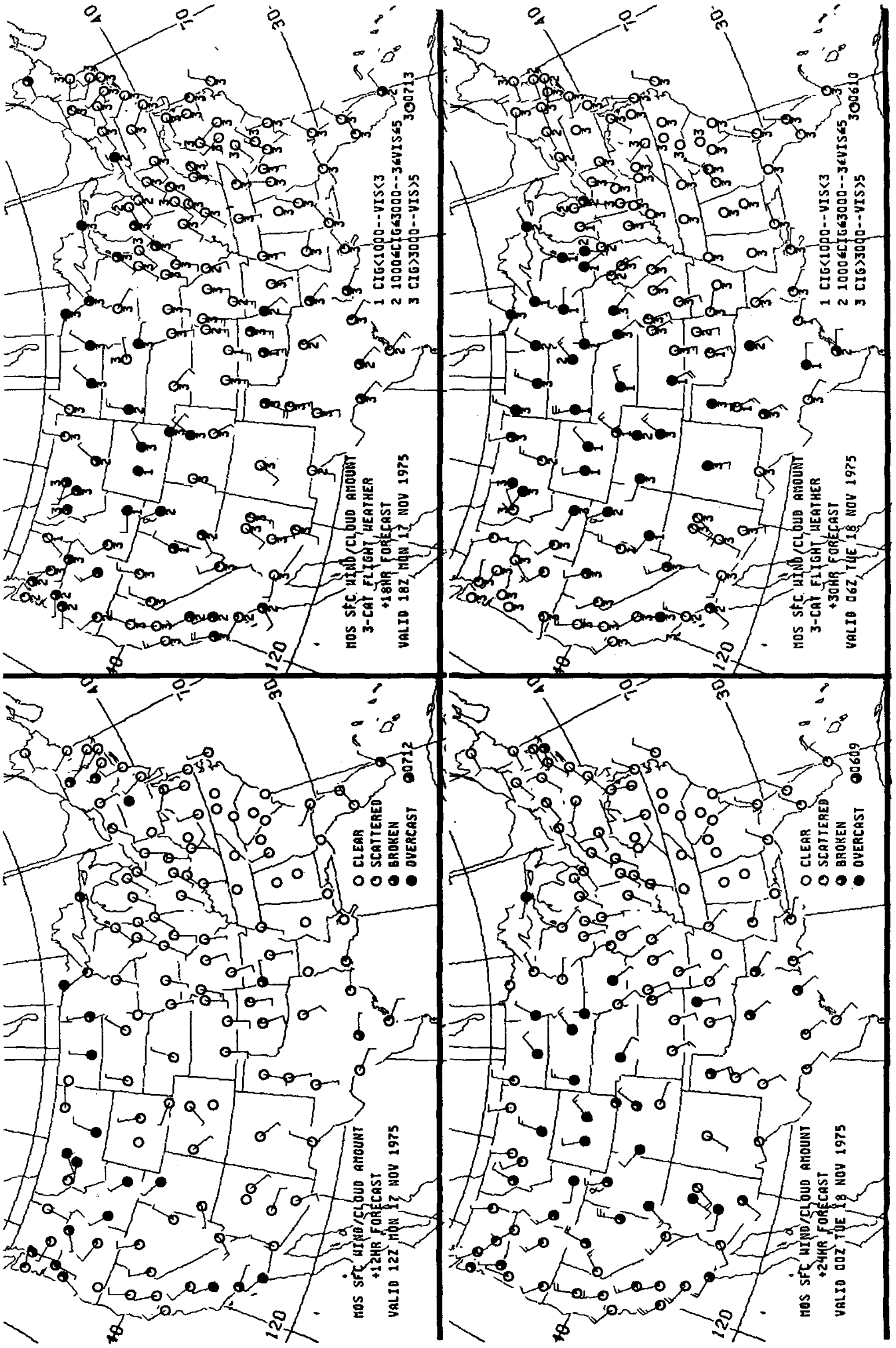

FIG. 1. Operational facsimile transmission of surface wind, cloud amount, and three-category flight weather forecasts for 12–30 h after 0000 GMT on 17 November 1975.

of the day of the year were also included. In addition for our 12 and 18 h forecast equations, we screened observed surface weather elements—sky cover, weather, *U* and *V* wind components, temperature and dew point—available 6 h after the input times for the numerical models.

The predictors from the forecast models were interpolated to the location of each of the 233 stations. To focus on local effects, we used data from only one station when we developed its equations. However, we space-smoothed some of the predictors over 5, 9 or 25 model grid points in order to reduce small-scale noise from the numerical model. The amount of smoothing increased with the projection time of the predictors.

The predictand data for these equations consisted of total sky cover observations categorized in the manner shown below.

Category	Total sky cover observation
1	Clear, thin scattered, thin broken, thin overcast
2	Scattered
3	Broken
4	Overcast, obscured

Thus, there were four predictands, each of which was binary, while predictors were expressed in both continuous and binary form. Specifically, a binary predictor was given a value of one if the variable from which it was derived was less than a particular threshold value; otherwise the value of the predictor was set to zero. This application of regression has been called REEP (Regression Estimation of Event Probabilities) by Miller (1964).

Using this approach, we derived separate equations for the 0000 and 1200 GMT runs of the PE and TJ models, and for each of seven projections—12, 18, 24, 30, 36, 42 and 48 h. We also derived backup equations without observations as predictors for the 12 and 18 h equations to handle situations where surface observations were missing. Furthermore, we developed an entire set of equations for the cool season and another for the warm season. Each cool season equation contained 10 predictors, while each warm season equation contained 12.

b. LFM-based forecasting equations

Later, also using the MOS technique, we derived 12-term regression equations for the 0000 and 1200 GMT runs of the LFM model. Separate equations were developed for each of 230 stations and projections of 6, 12, 18 and 24 h. This time, developmental data were from the warm seasons of 1973, 1974 and 1975.

Potential predictor fields from the LFM model were similar to those we screened for the PE- and TJ-based equations; except we added two dynamic predictors—relative vorticity and divergence—at differing projection times and levels throughout the atmosphere. In order to preserve the fine-scale detail of the LFM, we did not smooth any of these forecast fields over more than 5 grid points. We also screened sky cover, *U* and *V* wind components, temperature and dew point observations available 3 h after the LFM model input time for the 6 and 12 h equations.

The developmental predictand data consisted of more applicable opaque sky cover observations instead of the total sky cover observations we used in the earlier derivation of the PE- and TJ-based equations. Previously, we had to use total sky cover because opaque sky cover observations were not available in our data base until November 1972. We categorized these opaque sky cover observations as shown below.

Category	Tenths of opaque sky cover
1	0–1
2	2–5
3	6–9
4	10 (includes obscured)

Table 1 shows the warm season equations for the 12 h forecast from 0000 GMT for Oklahoma City. Column 1 gives the selected predictors and columns 5, 6, 7 and 8 show the coefficients, initial constants and the final reductions of variance for each of the four categories, which correspond roughly to clear, scattered, broken and overcast. Here, measures of moisture and stability, plus the observed sky cover, were important. The equations for all four categories have the same 12 predictors. This insures that the four probability estimates always sum to unity.

The low reductions of variance in Table 1 for categories 2 and 3 (i.e., scattered and broken) indicate that these categories are very difficult to forecast at Oklahoma City; this is also true for the rest of the stations.

3. Testing the forecasting equations

In order to evaluate the PE- and TJ-based system of estimating cloud amount, we compared experimental objective forecasts with NWS subjective local forecasts. Both sets of cloud forecasts were made for 92 widely scattered test stations during February and March 1974. These stations are also used in the National Weather Service's combined aviation/public weather verification system (National Weather Service, 1973). The objective forecasts were produced solely for verification purposes and had not been available as guidance to the field forecasters. The local forecasts were also recorded for verification purposes under instructions that the value recorded be ". . . not inconsistent with . . ." the official public weather forecasts.

Since our verification was primarily in terms of percent correct, we attempted to transform the objective probability estimates in a manner that would maximize this score. Initially, we did this by choosing the category with the highest probability as the categorical forecast.

The local forecasts and the observed data were both in tenths of opaque sky cover. We transformed these into categories in the same manner as shown in Section 2b.

Four-category, forecast-observed contingency tables were prepared using these transformed forecasts and opaque sky cover observations. Percent correct, skill score and bias by category (i.e., the number of forecasts in a particular category divided by the number of observations in that category) were computed from these tables. All the forecasts were valid 18 h from 0000 GMT. This valid time corresponds to an 18 h projection for the objective forecasts (i.e., observed surface predictors were not used for this test), and approximately a 9 h projection for the local forecasts. Table 2 shows the results by NWS Region as well as for all 92 stations combined.

The overall verification scores in Table 2 show that the objective forecasts (Objective A) did well in regard to percent correct—49% as compared with 48% for the local forecasts. The skill scores were also comparable—0.29 and 0.31, respectively. However, the bias by category scores for the local forecasts was much better (i.e., closer to unity) than that for the objective technique.

In order to correct this problem in bias, we devised a transformation that varied with each station and also was a function of the predictability of each category. This transformation involved using an inflation technique similar to the one proposed by Enger and described by Klein *et al.* (1959), plus the minimum bias matrix in Table 3.

Inflation insures that the variability, measured by the variance or standard deviation, of the observed and predicted values is nearly the same. This transformation can be written

$$\hat{P}'_j = \frac{\hat{P}_j - \bar{P}_j}{R_j} + \bar{P}_j, \qquad (1)$$

where $\hat{P}_j'$ is the "inflated" probability estimate for the jth category of cloud amount ($j=1,\ldots,4$), $\hat{P}_j$ the original objective probability forecast, $\bar{P}_j$ the average frequency of the cloud amount predictand from the developmental data sample, and R_j the multiple correlation coefficient of the predictand with the predictors in the forecasting equation.

TABLE 1. Equations for estimating four categories of cloud amount—clear, scattered, broken and overcast—12 h after 0000 GMT at Oklahoma City. The developmental sample consisted of 450 days from the warm seasons of 1973, 1974 and 1975.

Predictor	Threshold value	Projection (h)	Smoothing (points)	Coefficients			
				CAT1	CAT2	CAT3	CAT4
1. LFM mean relative humidity (720 to 490 mb)	Continuous	1200	None	−0.005	0.005	0.001	−0.001
2. LFM temperature difference (850–1000 mb)	Continuous	1200	None	−0.039	−0.016	0.003	0.053
3. Observed sky cover	Continuous	0300	None	−0.036	−0.004	0.026	0.014
4. LFM mean relative humidity (surface to 490 mb)	50%	1800	5	−0.003	0.125	0.043	−0.164
5. LFM temperature difference (700–850 mb)	Continuous	1200	None	0.063	−0.023	−0.025	−0.015
6. LFM relative vorticity $\times 10^5$ (850 mb)	Continuous	1200	None	−0.018	−0.034	0.009	0.043
7. LFM temperature difference (850–1000 mb)	−10K	1200	None	0.090	−0.173	−0.086	0.168
8. LFM mean relative humidity (surface to 490 mb)	40%	1200	None	0.188	−0.086	−0.135	0.033
9. LFM mean relative humidity (720 to 490 mb)	70%	1200	None	−0.084	0.165	0.116	−0.197
10. LFM mean relative humidity (surface to 490 mb)	Continuous	0600	None	0.001	−0.0001	−0.005	0.004
11. LFM mean relative humidity (surface to 490 mb)	60%	1800	5	0.203	0.014	−0.166	−0.051
12. LFM precipitable water	35 kg m^{-2}	1200	None	−0.048	−0.092	−0.036	0.176
	Initial constants			0.908	−0.422	0.173	0.342
	Total reductions of variance			0.410	0.096	0.111	0.361

Table 2. Comparative verification scores for subjective local and objective estimates of four categories of cloud amount—clear, scattered, broken and overcast—at 92 stations during February and March 1974. All forecasts were valid at 1800 GMT. The objective forecasts were based on the mode of the original four-category probability distribution (Objective A), and an inflated probability distribution adjusted by the minimum bias matrix in Table 3 (Objective B).

National Weather Service Regions	Type of forecast	Bias = number of forecasts/ number of observations				Percent correct	Skill score	Number of cases
		CAT1*	CAT2*	CAT3*	CAT4*			
Eastern	Local	0.63	1.47	1.18	0.86	51	0.338	1217
	Objective A	0.96	0.14	0.74	1.63	52	0.296	
	Objective B	0.78	1.08	0.83	1.17	53	0.340	
		(269)	(251)	(238)	(459)			
Southern	Local	0.72	1.86	1.16	0.58	48	0.301	1207
	Objective A	1.25	0.13	0.83	1.49	52	0.324	
	Objective B	1.06	1.08	0.87	0.94	51	0.328	
		(446)	(242)	(233)	(286)			
Central	Local	0.53	1.44	1.41	0.86	44	0.258	1474
	Objective A	1.14	0.01	1.07	1.57	46	0.251	
	Objective B	0.93	0.88	1.22	1.03	44	0.251	
		(413)	(342)	(261)	(458)			
Western	Local	0.91	1.00	1.09	1.02	51	0.339	842
	Objective A	1.04	0.01	0.75	1.85	47	0.272	
	Objective B	0.85	0.83	0.86	1.36	45	0.260	
		(232)	(178)	(183)	(249)			
Overall average	Local	0.68	1.47	1.23	0.83	48	0.309	4740
	Objective A	1.12	0.07	0.86	1.62	49	0.294	
	Objective B	0.93	0.97	0.96	1.11	48	0.302	
		(1360)	(1013)	(915)	(1452)			

* Number of observations are shown in parentheses.

After inflating all of the four-category probability forecasts, we multiplied them by the values in Table 3 as follows:

$$\hat{P}''_k=\sum_{j=1}^{4} a_{jk}\hat{P}'_j, \tag{2}$$

where a_{jk} is the weighting factor from column j and row k ($k=1,\ldots,4$) of the minimum bias matrix. The category with the largest value of $\hat{P}''_j$ was selected as the categorical forecast.

The forecasts labeled Objective B in Table 2 are the ones based on the inflation-minimum bias transformation. The minimum bias matrix shown in Table 3 was empirically formulated from the developmental data for this particular experiment. The overall skill scores—0.30 for the objective and 0.31 for the local forecasts—were very close, and the percent correct was 48% for both sets of forecasts. Furthermore, the bias by category values are now much better.

Encouraged by these test results, we developed two minimum bias matrices to be used with the cool season equations. One matrix was for cloud amount forecasts valid at 0600 and 1200 GMT, while the other was for the forecasts valid at 1800 and 0000 GMT. We used two matrices rather than one in an attempt to compensate for diurnal variations in cloudiness. These matrices are shown in Tables 4 and 5, respectively. We formulated the operational matrices from forecasts

Table 3. Experimental minimum bias matrix for making categorical forecasts of cloud amount for February and March 1974.

Observed category	Forecast category			
	1	2	3	4
1	120	55	0	0
2	60	100	40	0
3	0	55	90	50
4	0	0	40	90
Totals	180	210	170	140

Table 4. Operational minimum bias matrix for making categorical forecasts of cloud amount valid at 0600 GMT and 1200 GMT.

Observed category	Forecast category			
	1	2	3	4
1	110	10	0	0
2	10	150	10	0
3	0	10	145	10
4	0	0	10	110
Totals	120	170	165	120

TABLE 5. Operational minimum bias matrix for making categorical forecasts of cloud amount valid at 1800 GMT and 0000 GMT.

Observed category	Forecast category 1	2	3	4
1	100	10	0	0
2	10	170	10	0
3	0	10	150	10
4	0	0	10	100
Totals	110	190	170	110

based on 0000 GMT developmental data for all 233 stations during the cool season months from November 1972 through March 1973 and October 1973 through March 1974. Both of the matrices in Tables 4 and 5 are heavily biased in favor of forecasting categories 2 and 3 (scattered and broken) like the experimental matrix in Table 3.

More recently, we also examined the bias characteristics of forecasts from both the PE- and TJ-based and LFM-based equations during the warm season. These experiments were conducted with developmental data from the warm seasons of 1973, 1974 and 1975. In general, we discovered that the inflation technique alone produced acceptable bias characteristics for both forecasting systems. This appeared to be closely related to the greater frequency of convectively induced, scattered and broken cloudiness during the warm season. Also, the use of opaque sky cover categories in the development of the LFM-based equations appeared to further improve the bias values for this system. Thus, we did not develop any minimum bias matrices for use with the warm season cloud equations.

4. The operational forecasting system

Our PE- and TJ-based system to forecast cloud amount first went into daily operation at the National Meteorological Center in December 1974. Probability estimates for clear, scattered, broken and overcast conditions, as well as the transformed categorical (i.e., "best" category) forecast, have been available since then on teletypewriter on a request basis through the Federal Aviation Administration's Weather Message Switching Center at Kansas City. These are now called "final" guidance cloud forecasts.

Since August 1975, final guidance categorical forecasts have also appeared on a four-panel (12, 18, 24 and 30 h) facsimile chart (see Fig. 1) on the NWS Forecast Office Facsimile Circuit. In addition, automated forecasts of surface wind (Carter, 1975a) and flight weather (combined categories of ceiling and visibility; National Weather Service, 1975) based on MOS are shown on this map. All of the forecasts are updated every 12 h.

In May 1976 the "early" guidance (LFM-based) cloud forecasts also became operational. These four-category probability estimates and single best category forecasts are available only on teletypewriter on a request basis.

5. Verification of operational forecasts

We verified operational final guidance (PE- and TJ-based) cloud forecasts for 88 stations for the period December 1974 through March 1975 in conjunction with the combined aviation/public weather verification system of the NWS. These best category guidance forecasts had been generated using the inflation-minimum bias technique. We transformed the local forecasts and opaque sky cover observations into categorical form in the same manner as described earlier (see Section 2b). The objective forecasts were for projections of 18, 30 and 42 h from 0000 GMT (i.e., observed surface predictors were not used). The local forecasts were issued about 1000 GMT and may have been based on data as much as 9 h later than those used in the guidance fore-

TABLE 6. Comparative verification scores for subjective local and objective guidance forecasts of four categories of cloud amount—clear, scattered, broken and overcast—at 88 stations during December 1974 through March 1975. The objective forecasts were made from the 0000 GMT runs of the PE and TJ models.

Projection (h)	Type of forecast	Bias = Number of forecasts/ number of observations CAT1*	CAT2*	CAT3*	CAT4*	Percent correct	Skill score	Number cases
18	Guidance	1.28	0.88	0.88	0.96	51	0.31	8124
	Local	0.66	1.44	1.35	0.83	50	0.32	
		(1823)	(1497)	(1532)	(3272)			
30	Guidance	1.23	0.80	0.81	0.93	57	0.35	8048
	Local	0.60	2.12	2.31	0.69	44	0.25	
		(2644)	(1002)	(811)	(3591)			
42	Guidance	1.31	0.86	0.80	0.98	47	0.26	8134
	Local	0.47	1.78	1.54	0.68	39	0.19	
		(1843)	(1513)	(1536)	(3242)			

* Number of observations are shown in parentheses.

TABLE 7. Comparative verification scores for subjective local and objective guidance forecasts of four categories of cloud amount—clear, scattered, broken and overcast—at 94 stations during April through September 1975. The objective forecasts were made from the 0000 GMT runs of the PE and TJ models.

Projection (h)	Type of forecast	Bias = Number of forecasts/ number of observations				Percent correct	Skill score	Number of cases
		CAT1*	CAT2*	CAT3*	CAT4*			
18	Guidance	0.87	1.07	1.10	0.98	44	0.25	15120
	Local	0.69	1.50	1.10	0.64	49	0.31	
		(4301)	(4123)	(3657)	(3039)			
30	Guidance	1.05	0.85	1.10	0.97	48	0.24	14975
	Local	0.68	1.87	1.73	0.54	43	0.23	
		(6685)	(2826)	(1821)	(3643)			
42	Guidance	0.93	1.11	1.06	0.86	41	0.21	15094
	Local	0.57	1.76	1.09	0.47	40	0.19	
		(4303)	(4150)	(3635)	(3006)			

* Number of observations are shown in parentheses.

casts. The verification scores for these subjective local and objective guidance forecasts are shown in Table 6.

The percent correct and skill scores in Table 6 indicate that the guidance forecasts were better than the local forecasts for the 30 and 42 h projections. Both sets of forecasts were about equal in skill for the shorter (18 h) projection. The bias results show that the guidance tended to overestimate the occurrence of clear skies (category 1), while the local forecasts consistently overestimated scattered and broken (categories 2 and 3). For more details, see Carter (1975b).

We also comparatively verified operational final guidance cloud forecasts for 94 stations for the period April through September 1975 (Carter, 1976). Unfortunately, due to implementation problems, the guidance forecasts had been based on cool season equations during the first four months of this period. Furthermore, the 18 h guidance forecasts did not begin using observed input until August 1975. Table 7 gives the verification scores for these local and guidance forecasts.

The scores in Table 7 show that the local estimates of cloudiness were substantially better than the guidance for the 18 h projection. However, the guidance was superior at 30 h. Both systems were nearly equal in overall skill for the 42 h forecasts. Again, the bias by category scores indicates that the local forecasts strongly overestimated scattered conditions, and to a lesser extent broken clouds.

6. Conclusions and plans

In general, the results of this study show that the MOS technique is useful in forecasting cloud amount. From our testing, it appears that a satisfactory way of obtaining nearly unbiased categorical forecasts (i.e., forecasting each category as often as it occurs) during the cool season is to transform the original probability estimates using both the inflation technique and minimum bias matrices. However, the inflation method alone appears to produce nearly unbiased forecasts during the warm season.

Comparisons of experimental final guidance (PE- and TJ-based) forecasts and National Weather Service local forecasts indicate that the objective estimates were generally as accurate as the subjective forecasts. We obtained similar results when we verified operational final guidance forecasts of cloud amount during both the cool and warm seasons.

We plan to verify operational early guidance (LFM-based) forecasts produced during the warm season of 1976. Early guidance equations for the cool season are being developed. We also plan to extend the final guidance forecasting system to Alaska and Hawaii.

Any substantial improvement in this objective cloud forecasting system will most likely result from the inclusion of predictors from new or improved numerical models and/or new methods for producing categorical forecasts of scattered and broken.

Acknowledgments. We are very grateful to Messrs. Frank Lewis, Fred Marshall and George Hollenbaugh for valuable support during the development and implementation of the forecast system described here. We also appreciate Mr. Joel Nathan's assistance in preparing the facsimile chart.

REFERENCES

Carter, G. M., 1975a: Automated prediction of surface wind from numerical model output. *Mon. Wea. Rev.*, **103**, 866–873.

——, 1975b: Comparative verification of local and guidance cloud amount forecasts—No. 1. Office Note 75-7, Techniques Development Laboratory, Silver Spring, Md., 8 pp.

——, 1976: Comparative verification of local and guidance cloud amount forecasts—No. 2. Office Note 76-8, Techniques Development Laboratory, Silver Spring, Md., 13 pp.

Glahn, H. R., 1974a: An objective cloud forecasting system. *Pre-*

prints Fifth Conf. Weather Forecasting and Analysis, St. Louis, Amer. Meteor. Soc., 79–80.

——, 1974b: The TDL MOS development system IBM 360/195 version. Office Note 74-14, Techniques Development Laboratory, Silver Spring, Md., 67 pp.

Howcroft, J., and A. Desmaris, 1971: The limited area fine mesh (LFM) model. *NWS Tech. Proc. Bull.*, No. 67, 11 pp.

Klein, W. H., and H. R. Glahn, 1974: Forecasting local weather by means of model output statistics. *Bull. Amer. Meteor. Soc.*, **55**, 1217–1227.

——, B. M. Lewis and I. Enger, 1959: Objective prediction of five day mean temperature during winter. *J. Appl. Meteor.*, **16**, 672–682.

Miller, R. G., 1964: Regression estimation of event probabilities. Tech. Rep. No. 1, Contract Cwb-10704, The Travelers Research Center, Inc., 153 pp.

National Weather Service, 1973: Combined aviation/public weather forecast verification. *Operations Manual*, C-73.

——, 1975: MOS surface wind, cloud amount, and 3-category flight weather forecasts. *NWS Tech. Proc. Bull.*, No. 139, 10 pp.

Shuman, F. G., and J. B. Hovermale, 1968: An operational six-layer primitive equation model. *J. Appl. Meteor.*, **7**, 525–547.

Reap, R. M., 1972: An operational three-dimensional trajectory model. *J. Appl. Meteor.*, **11**, 1193–1202.